Machine learning constraints on deviations from general relativity from the large scale structure of the Universe

George Alestas[1,2,*] Lavrentios Kazantzidis,[2,†] and Savvas Nesseris[1,‡]

[1]*Instituto de Física Teórica UAM-CSIC, Universidad Autonóma de Madrid, Cantoblanco, 28049 Madrid, Spain*
[2]*Department of Physics, University of Ioannina, GR-45110 Ioannina, Greece*

We use a particular machine learning approach, called the genetic algorithms (GAs), in order to place constraints on deviations from general relativity (GR) via a possible evolution of Newton's constant $\mu \equiv G_{\text{eff}}/G_{\text{N}}$ and of the dark energy anisotropic stress η, both defined to be equal to one in GR. Specifically, we use a plethora of background and linear-order perturbations data, such as Type Ia supernovae, baryon acoustic oscillations, cosmic chronometers, redshift space distortions, and E_g data. We find that although the GA is affected by the lower quality of the currently available data, especially from the E_g data, the reconstruction of Newton's constant is consistent with a constant value within the errors. On the other hand, the anisotropic stress deviates strongly from unity due to the sparsity and the systematics of the E_g data. Finally, we also create synthetic data based on a next-generation survey and forecast the limits of any possible detection of deviations from GR. In particular, we use two fiducial models: one based on the cosmological constant ΛCDM model and another on a model with an evolving Newton's constant, dubbed μCDM. We find that the GA reconstructions of $\mu(z)$ and $\eta(z)$ can be constrained to within a few percent of the fiducial models and in the case of the μCDM mocks, they can also provide a strong detection of several σs, thus demonstrating the utility of the GA reconstruction approach.

I. INTRODUCTION

Currently the de facto cosmological model is widely considered to be the Λ cold dark matter (ΛCDM) scenario, which is based on general relativity (GR) and it contains not only ordinary matter but also incorporates two other ingredients that are the backbone of this model. The cosmological constant Λ, is crucial for the description of the accelerated expansion of the Universe, as well as the CDM. Ever since the observed accelerated expansion of the Universe it holds, almost unanimously, the scepters in the field of cosmology. Lately, however, there is mounting evidence supporting the idea that perhaps ΛCDM is not infallible.

To that end, there have been observed a number of discrepancies/tensions (for comprehensive reviews see Refs. [1–3]) that emerge when comparing the predictions made by the model to observational data, such as the S_8 tension [4–11], cosmic microwave background (CMB) anisotropy anomalies [12–17], the lithium problem [18–23], etc. Arguably though, the most well known of these tensions is the so-called Hubble tension [1–3,24–29]. This tension

involves the 5σ disagreement between the value of H_0 as given by the Planck collaboration via the CMB data ($H_0 = 67.27 \pm 0.60$ km s^{-1} Mpc^{-1}) [30], and the one given by the SH0ES team using SNIa data calibrated by Cepheids ($H_0 = 73.04 \pm 1.04$ km s^{-1} Mpc^{-1}) [31].

If these tensions are not due to some unknown systematic effect, then some of them might suggest that ΛCDM needs to be further modified. As a result, a number of possible early or/and late time solutions to these tensions have been proposed in the literature, including a variety of modified gravity models. A number of recent works [32–38] in the field argue that in order to reconcile the Hubble and growth tensions, an evolution of Newton's constant G_{eff} that leads to its decrease as the Universe evolves is needed. More specifically, this modified gravity solution involves ultra-late time gravitational transitions that, in their core, explore and enable the possibility of a transition in the Type Ia supernova (SNIa) absolute magnitude M_B, which would allow for the easing of both the Hubble and S_8 tensions, whilst leaving the standard cosmology intact within the ultralow redshift range $z \in [0.01, 1000]$.

This transition in M_B could very well be a product of a gravitational transition of G_{eff}, since the two are connected via the Chandrasekhar mass m_{ch} that evolves as $m_{\text{ch}} \sim G_{\text{eff}}^{-3/2}$, leading to $\Delta M_B = \frac{15}{4} \log \mu$ [39,40], where

*g.alestas@csic.es
†l.c.kazantzidis@gmail.com
‡savvas.nesseris@csic.es

$$\mu \equiv \frac{G_{\text{eff}}}{G_{\text{N}}}, \tag{1}$$

and G_{N} is the bare Newton's constant as measured by Cavendish-type experiments in a laboratory.

Within the context of this paper, we use machine learning (ML) in the form of genetic algorithms (GAs) in order to assess whether current cosmological data from large scale structure (LSS) probes predict an evolution of the G_{eff}. The advantage of ML methods, in this case, is that they allow for a bottom-up reconstruction of G_{eff} solely from the data, avoiding theoretical biases and enabling us to look for features that may otherwise not be readily detected via more traditional approaches [41].

More specifically, we use the Pantheon sample of SNIa data, an up-to-date collection of baryon acoustic oscillations (BAO) data, cosmic chronometer (CC) data, an up-to-date compilation of growth rate/redshift space distortions data, and a similarly up-to-date collection of E_g data. Via the GA process, we are able to ascertain the best-fit functions for the luminosity distance $D_L(z)$ using the Pantheon SNIa data, the angular diameter distance $D_A(z)$ using the BAO data, the Hubble parameter $H(z)$ using the CC data, the growth rate of perturbations $f\sigma_8(z)$ via the growth rate data and the $P_2(z)$ observable, which is a model-independent probe of modified gravity defined as the ratio of the lensing function Σ multiplied by $\Omega_{\text{m},0}$ and the growth rate f [42,43], using the E_g data. The combination of these cosmological quantities allows us to probe the possible evolution of Eq. (1).

In linear order perturbation theory one more quantity can also be defined in order to parametrize deviations from GR, namely the anisotropic stress η. The latter is equal to the ratio of the Newtonian potentials and, in the absence of any matter-induced anisotropic stresses, is equal to unity. Here, we also use the GA in order to reconstruct η and we compare those results with an analysis using earlier data by Ref. [41].

Having done the GA reconstructions with the currently available data, we also forecast the constraining power of approach using mock data based on a future next-generation survey like Euclid or DESI, albeit without targeting any of them specifically, in order to determine how accurately forthcoming surveys will be able to constrain deviations from GR. To do so, we use two mocks, one based on ΛCDM and one more, using parametrizations for $\mu(z)$ and $\eta(z)$.

The structure of our paper is as follows: in Sec. III we present a brief summary of the GA approach and how it works, in Sec. IV we present the compilations of currently available and mock data we use, while in Sec. V we present the results of our analysis in the cases of both the currently available and mock data. Finally, in Sec. VI we present our conclusions.

II. THEORY

Here we describe in more detail the theoretical setup of our analysis, which is based on cosmological perturbation theory, setting the stage for the machine learning reconstructions in later sections. We mainly follow Refs. [10,41,44], albeit with slight changes in the notation.

First, we assume a perturbed flat Friedmann-Lemaître-Robertson-Walker (FLRW) metric

$$ds^2 = -(1 + 2\Psi)dt^2 + a(t)^2(1 - 2\Phi)d\vec{x}^2, \tag{2}$$

where $a(t)$ is the scale factor given in terms of the cosmic time t, while Φ and Ψ are the scalar potentials in the Newtonian gauge. Using the field equations it is possible to show that, for a plethora of different modified gravity models, in the subhorizon and quasistatic regime, the Newtonian potentials obey the following Poisson equations in Fourier space [45]

$$-\frac{k^2}{a^2}\Psi = 4\pi G_{\text{N}}\mu(k,a)\rho_m\delta_m, \tag{3}$$

$$-\frac{k^2}{a^2}(\Psi + \Phi) = 4\pi G_{\text{N}}\Sigma(k,a)\rho_m\delta_m, \tag{4}$$

where k is the wave number, ρ_m is the matter density, $\delta_m \equiv \frac{\delta\rho_m}{\rho_m}$ is the growth factor of matter density perturbations and the parameters Σ and μ characterize deviations from GR. In particular, in GR the corresponding limits are $\Sigma = 2$ and $\mu = 1$, see Refs. [41,44,46], while in other theories they can be in a general scale and time dependent [45].

Furthermore, modified gravity models in general may also induce an anisotropic stress, which can be parametrized by the parameter η defined as the ratio of the two potentials

$$\eta \equiv \frac{\Phi}{\Psi}, \tag{5}$$

which in GR has the limit $\eta = 1$ as expected in the absence of anisotropic stresses from matter (e.g., from neutrinos or photons [46]). The anisotropic stress or gravitational slip quantity $\eta(z)$ is both an element that allows for the nonminimal coupling of the dark energy with gravity in the context of the Jordan frame, and a purely geometric characteristic of higher order modified gravity models [47–49].

The evolution of the cosmological overdensity $\delta_m(a)$ satisfies, in the absence of neutrinos and for most modified gravity theories, the following differential equation in the subhorizon regime

$$\delta_m''(a) + \left(\frac{3}{a} + \frac{H'(a)}{H(a)}\right)\delta_m'(a) - \frac{3}{2}\frac{\Omega_{\text{m},0}\mu(k,a)}{a^5 H^2(a)/H_0^2}\delta_m(a) = 0, \tag{6}$$

where primes denote differentiation with respect to a and k corresponds to the scale. Observations strongly suggest the existence of a large-scale structure of the Universe that was created during a matter domination era, typically assumed to be $z \in [1, 10^4]$. It is then easy to show that the growth should behave as $\delta(a \ll 1) \sim a$ at an initial time deep in matter domination.

A related quantity is the growth rate f, which is defined as

$$f(a) \equiv \frac{\mathrm{d} \ln \delta_m(a)}{\mathrm{d} \ln a}, \qquad (7)$$

and is a proxy for the growth of matter density perturbations on large scales. However, in the past two decades the vast majority of LSS surveys report instead the bias-independent product $f\sigma_8(a) = f(a) \cdot \sigma_8(a)$, where

$$\sigma_8(a) \equiv \frac{\sigma_8}{\delta_m(1)} \delta_m(a), \qquad (8)$$

with σ_8 corresponding to the density rms fluctuations within spheres of radius on scales of about $8h^{-1}$ Mpc.

Therefore, given a Hubble rate $H(a)$ and a parametrization for $\mu(k, a)$, Eq. (6) can be solved either numerically or analytically and using the solution, the theoretical prediction for $f\sigma_8$ is readily constructed via Eqs. (7) and (8).

In order to reconstruct the aforementioned parameters μ, Σ, and η, we follow Ref. [44] and we define a set of variables

$$P_2(a) = \frac{\Sigma \Omega_{\mathrm{m},0}}{f}, \qquad (9)$$

$$P_3(a) = \frac{\mathrm{d} \ln f\sigma_8}{\mathrm{d} \ln a}, \qquad (10)$$

$$E(a) = H(a)/H_0, \qquad (11)$$

where $\Omega_{\mathrm{m},0}$ is the present value of the matter density parameter. Then, it follows that the anisotropic stress is given by the relation [44]

$$1 + \eta(a) = \frac{3P_2(a)a^{-3}}{2E(a)^2 \left[P_3(a) + 2 + \frac{\mathrm{d} \ln E}{\mathrm{d} \ln a}\right]}. \qquad (12)$$

Combining Eqs. (3) and (4) and (5) we can find the expression

$$\Sigma = \mu(1 + \eta), \qquad (13)$$

which then from the definition of P_2 via Eq. (9), gives

$$\mu = \frac{fP_2}{(1 + \eta)\Omega_{\mathrm{m},0}}. \qquad (14)$$

In the latter equation the final ingredient is the parameter P_2, which can be shown to be related to the so-called E_g statistic, see Ref. [44] and references therein. In fact, E_g is the expectation value of the ratio of lensing and galaxy clustering observables at a scale k

$$E_g = \left\langle \frac{a\nabla^2(\Psi + \Phi)}{3H_0 f \delta_m} \right\rangle_k, \qquad (15)$$

which after using the Poisson equation (3) and (4) reduces to $E_g = 2P_2$. As $\Omega_{\mathrm{m},0}$ is not directly observable but can only be inferred by other probes, we can rewrite Eqs. (12) and (14) in terms of the actual observables as

$$\Omega_{\mathrm{m},0}\mu(a) = \frac{2f(a)E_g(a)}{1 + \eta(a)}, \qquad (16)$$

$$1 + \eta(a) = \frac{3E_g(a)a^{-3}}{E(a)^2 \left[P_3(a) + 2 + \frac{\mathrm{d} \ln E}{\mathrm{d} \ln a}\right]}, \qquad (17)$$

where in what follows we will consider the combination $\Omega_{\mathrm{m},0}\mu(a)$ as it is independent from $\Omega_{\mathrm{m},0}$. Still, any deviation of Eq. (16) at any redshift from a constant value would be a smoking-gun signature for deviations from GR at late times [50], so this makes it an extremely useful statistic.

III. THE GENETIC ALGORITHMS

In this section we describe the GA, as used in our work. The GA has had several applications in cosmology, see for example Refs. [41,51–57], in forecasts of future LSS surveys [58–60], but also in a wide range of areas such as particle physics [61–63], astronomy and astrophysics [64–66] and other fields like computational science, economics, medicine and engineering [67,68]. There are also several other similar symbolic regression methods, see for example [69–76].

The GA that we use in the present work in order to identify a possible evolution of G_{eff} correspond to an unsupervised symbolic regression of data process that in reality simulates the biological natural selection. Initially, a set of randomly selected orthogonal basis functions (which we denote as "grammar") are chosen and are then subjected to crossover and mutation operations over time, until a specific termination criterion is met. For a set of vectors (or functions), orthogonality is a much stronger statement as it implies not only they are linearly independent, but also their dot product is zero. This is necessary in order to avoid degeneracies between the functions that could otherwise affect the convergence rate of the algorithm.

In particular, the initial "grammar" population corresponds to the first generation of functions, in which we need to impose any appropriate conditions in order to obtain a physically meaningful result and applying the maximum likelihood method, a χ^2 value is calculated for each member. Then, convergence is achieved when the χ^2 does not improve or change for a few hundred generations. As can be seen in Fig. 1 of Ref. [57], when this happens it typically implies that the algorithm is at the minimum and it stays there, unlike an MCMC where the GA random walker of a particular chain "oscillates" around it.

Next, the best-fit functions from the first generation are determined through a tournament selection method, i.e., a method that randomly chooses a subgroup of candidates and picks the dominant one from each subgroup. After that, the second generation of candidate functions is obtained by sequentially applying the crossover and mutation operations. Then this process is repeated a thousand times in order to guarantee convergence. Notice that with the term "crossover" we refer to the combination of two parties in order to form a descendant, while with the term mutation we describe the arbitrary change of an individual. A flowchart of a typical run of a GA can be seen in Fig. 1.

Naively one could think that grammar selection is crucial in order to obtain any physically meaningful results, however as it has been discussed in Ref. [51] the initial population of the grammar is irrelevant to the outcome and only influences the convergence rate. Regarding the termination criterion as it has been extensively discussed in Ref. [51] a number of different conditions can be applied. In this paper we let the GA reach the maximum number of generations and choose the best candidate based on a normal χ^2 statistic.

As we have already mentioned one of the advantages of such a method is its model-independent nonparametric nature. As a result, the final outcome of the process corresponds to a group of continuous functions with respect to a variable x (in our case we choose the variable to be the redshift z) and not to specific best-fit parameters of a model. However, this advantage can also be considered as a simultaneous drawback of the method since the lack of parameters leads to the abandonment of standard techniques for the calculation/propagation of the errors. Fortunately, this shortcoming can be bypassed through the path integral approach, as discussed in Ref. [53].

In a nutshell, as the GA explores the whole functional space, it effectively (in principle) passes through every possible value of the reconstructed value $f(x)$ at each x. This is effectively the same as doing a path integral and normalizing the likelihood reduces to a set of Gaussian integrals which result to a sum of error functions, see Ref. [53]. This approach at determining the errors from the GA has been well-tested against bootstrap Monte Carlo and fisher matrix techniques and has been found to be in good agreement. Furthermore, it has the added advantage that it is much faster compared to bootstrap and requires less computational power.

IV. THE DATA

As we discussed in Sec. I, in order to constrain a possible evolution of the evolving Newton's constant G_{eff} through the GA, we use up-to-date cosmological data compilations, including SNIa, BAO, and CC points, as well as growth and E_g data. Note, we do not use any CMB data (even though it would have decreased the errors) as that would leave a huge gap in the points between $z \simeq 2$ and $z \simeq 1090$ that we would have to interpolate in between. The exact compilations and the details for the likelihoods are described in what follows.

A. Type Ia supernovae: The Pantheon sample

The SNIa dataset that we use in this work is the Pantheon sample [78]. This sample corresponds to one of the largest compilations of SNIa, which includes data both in the high and low redshift regimes that are observed by six different surveys. As a result, the sample incorporates 1048 SNeIa in the redshift range $z \in (0.01, 2.26)$.[1] In general, SNIa are used extensively in the literature in order to measure the expansion rate $H(z)$ since they can be directly connected with the theoretically predicted apparent magnitude of a model $m_{th}(z)$ through the following equation:

$$m_{\text{th}}(z) = M_B + 5\log_{10}\left[\frac{D_L(z)}{\text{Mpc}}\right] + 25, \qquad (18)$$

with M_B corresponding to the corrected absolute magnitude and D_L is the luminosity distance, which in a flat FLRW Universe is given by
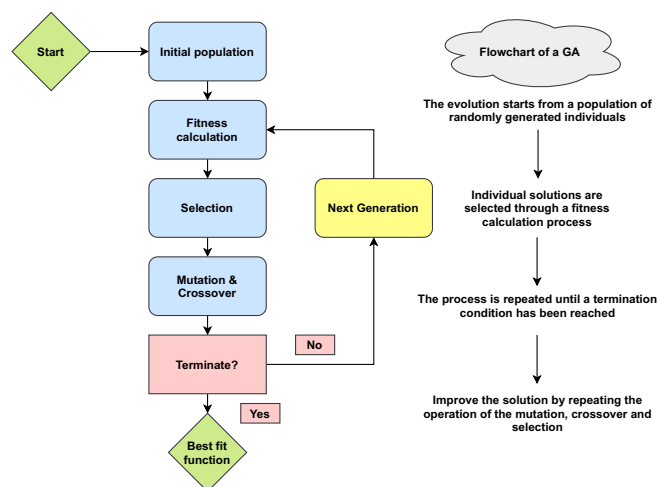

FIG. 1. Flowchart of a typical run of a genetic algorithm (from Ref. [77]).

---

[1]The newer compilation of the Pantheon+ was made publicly available in the final stages of this work, but we have not included it as we do not expect it to significantly affect our analysis.

$$D_L(z) = \frac{c}{H_0}(1+z)\int_0^z \frac{1}{E(z')}\,dz'. \qquad (19)$$

Even though some recent works [79–83] argue that the degenerate combination of this quantity with H_0 that appears when Eq. (18) is expressed using the Hubble free luminosity distance $D_L(z)$ may contain useful information, normally the absolute magnitude M_B is treated as a nuisance parameter and is marginalized [78,84,85] following the method discussed in the Appendix C of Ref. [85]. So, in the present work we adopt the standard method and marginalize over M_B, leading as a result to a χ^2 function of the form [41]

$$\chi^2_{\text{SNIa}} = A - \frac{B^2}{E} + \ln\left(\frac{E}{2\pi}\right), \qquad (20)$$

where

$$A = \Delta\vec{m}\cdot\vec{C}^{-1}\cdot\Delta\vec{m},$$
$$B = \Delta\vec{m}\cdot\vec{C}^{-1}\cdot\Delta\vec{I},$$
$$E = \vec{I}\cdot\vec{C}^{-1}\cdot\vec{I}.$$

In the above expression $\vec{C}$ is the total covariance matrix (including both the statistical and systematic uncertainties), $\vec{I} = (1, 1, \ldots, 1)$ and $\Delta m = m_{\text{obs}}(z_i) - m_{\text{th}}(z)$.

The total covariance matrix $\vec{C}$ arises as the sum of two independent matrices [78]. The first one is a diagonal matrix that includes the statistical uncertainties of the apparent magnitudes m_{obs} of each SNIa that has the following form:

$$\begin{pmatrix} \sigma^2_{m_{\text{obs,1}}} & 0 & \ldots & 0 \\ 0 & \sigma^2_{m_{\text{obs,2}}} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma^2_{m_{\text{obs,1048}}} \end{pmatrix}, \qquad (21)$$

while the second one is a nondiagonal matrix related to systematic uncertainties. Therefore, for a specific form for the Hubble rate $H(z)$, the theoretically predicted apparent magnitude can be produced using Eq. (18) and then the quality of the fit can be estimated applying the maximum likelihood method, i.e., minimizing Eq. (20).

B. Baryon acoustic oscillations

Regarding the BAO compilation used in the present analysis we employ data from different observational missions such as the 6dFGS [86], WiggleZ [87,88], DES [89], and Lya [90] surveys as well as other type of BAO data from the fourth generation of the SDSS mission

(SDSS-IV), i.e., the extended baryon oscillation spectroscopic survey (eBOSS) [91]. However, since we incorporate data from different missions we cannot construct a unique χ^2 formula as we did in the SNIa case, because the different missions do not constrain the same observational quantities. In reality, they report different dependent combinations that are related to one another. So, before we establish the corresponding χ^2 formula, we need to recall some basic definitions.

It is known that the BAO correspond to a relic from the prerecombination era of the Universe that imposes a characteristic scale occurring either as a peak in the galaxy correlation function or as damped oscillations in the power spectrum. The radius of the sound horizon at last scattering r_s is defined as [92]

$$r_s(z_d) \equiv \int_{z_d}^\infty \frac{c_s(z)}{H(z)}\,dz, \qquad (22)$$

where the lower limit of the integration corresponds to the drag redshift z_d, i.e., a redshift shortly after recombination that can be either calculated through a numerical package such as CAMB or through the approximate formula discussed in Ref. [93] [see its Eq. (4)]. The numerator of (22) describes the sound speed of the baryonic-photon fluid given as [94]

$$c_s(z) = \frac{c}{\sqrt{3\left(1 + \frac{3\Omega_{b,0}}{4\Omega_{\gamma,0}}\frac{1}{1+z}\right)}}, \qquad (23)$$

where $\Omega_{b,0}$ and $\Omega_{\gamma,0}$ are the present day baryon and photon densities, respectively.

Taking into account the physical angular diameter distance $D_A(z)$ that in the context of a flat Universe is given as

$$D_A(z) = \frac{c}{1+z}\int_0^z \frac{dz'}{H(z')}, \qquad (24)$$

two different observational quantities that are widely used in the literature can be obtained. The first one corresponds to the comoving angular diameter distance D_M that is given through the following simple formula [95]:

$$D_M(z) = (1+z)D_A(z), \qquad (25)$$

while the second one is the combination

$$D_V(z) = \left[\frac{czD_M^2(z)}{H(z)}\right]^{1/3}. \qquad (26)$$

Even though, Eqs. (25) and (26) are the main observational quantities that are published by the BAO surveys, some surveys may also report the D_H observable that is defined as

$$D_H(z) \equiv \frac{c}{H(z)}. \tag{27}$$

Now returning to the specific data compilation of the current work, we use the results of the following surveys:

(i) The 6dFGS and WiggleZ missions that report the quantity $d_z \equiv r_s(z_d)/D_V(z)$ giving the following values

Missions	z	d_z	σ_{d_z}	References	Date
6dFGS	0.106	0.336	0.015	[86]	June 2011
WiggleZ	0.44	0.073	0.031	[87,88]	April 2012
	0.6	0.0726	0.0164		
	0.73	0.0592	0.0185		

which lead to a χ^2 formula of the form

$$\chi^2_{\text{6dFS,WigZ}} = V^i C_{ij}^{-1} V^j. \tag{28}$$

In this case, the vector V^i is given as $V^i = d_{z,i} - d_z(z_i)$, whereas the covariance matrix C_{ij} yields

$$\begin{pmatrix} \frac{1}{0.015^2} & 0 & 0 & 0 \\ 0 & 1040.3 & -807.5 & 336.8 \\ 0 & -807.5 & 3720.3 & -1551.9 \\ 0 & 336.8 & -1551.9 & 2914.9 \end{pmatrix}. \tag{29}$$

(ii) The DES mission, that publishes the combination D_M/r_s to be

Mission	z	D_M/r_s	σ_{D_M/r_s}	References	Date
DES	0.835	18.92	0.51	[89]	July 2021

which gives a χ^2 formula of the form

$$\chi^2_{\text{DES}} = \sum_i \left(\frac{D_M(z,i)/r_s - D_M(z_i)/r_s}{\sigma_{D_M(z,i)/r_s}} \right)^2. \tag{30}$$

(iii) The Lya mission, which constrains the combinations $f_{\text{BAO}} = (D_H/r_s, D_M/r_s)$ as

Mission	z	f_{BAO}	$\sigma_{f_{\text{BAO}}}$	References	Date
Lya	2.33	(8.99, 37.5)	(0.19, 1.1)	[90]	July 2020

Notice that the 1σ errors provided in the above table correspond to the statistical errors. So taking into account the two autocorrelations along with the two cross-correlations as illustrated in Eq. (43) of Ref. [90] the relevant χ^2 function is constructed as

$$\chi^2_{\text{Lya}} = V^i C_{ij}^{-1} V^j. \tag{31}$$

In this case, the vector V^i reads as $V^i = (f_{\text{BAO}}(z, i) - f_{\text{BAO}}(z_i)) \equiv (D_H(z, i)/r_s - D_H(z_i)/r_s, D_M(z, i)/r_s - D_M(z_i)/r_s)$.

(iv) Finally we include the results of the eBOSS mission which contains measurements in various redshift ranges as summarized in detail in Table 3 of Ref. [91]. Using these values, a χ^2 formula for each data point can be constructed of the form

$$\chi^2_{\text{eBOSS}} = V^i C_{ij}^{-1} V^j, \tag{32}$$

where the vector V^i corresponds to the difference between the observational values with the respected theoretical expressions and C_{ij} is the corresponding covariance matrix.

Note that some of the individual χ^2 terms might be correlated with each other, e.g., the Lya terms and the ones from eBOSS. Unfortunately however, it is impossible to calculate the possible correlations between the terms as we do not have the necessary covariance matrices, so we will assume they are uncorrelated. Therefore, to obtain the total χ^2, we add all the individual terms as follows [41]:

$$\chi^2_{\text{BAO}}(r_s h, \text{GA}) = \chi^2_{\text{6dFS,WigZ}} + \chi^2_{\text{DES}} + \chi^2_{\text{Lya}} + \chi^2_{\text{eBOSS}}. \tag{33}$$

Furthermore, as mentioned earlier the BAO points depend strongly on the sound horizon at the drag epoch $r_s(z_d)$ times the dimensionless Hubble parameter $h = H_0/(100 \text{ km s}^{-1} \text{Mpc}^{-1})$, i.e., the combination $r_s h$. Schematically, this means that the χ^2 is a function of the product of the two variables and the functions produced by the GA, i.e., $\chi^2 = \chi^2(r_s h, \text{GA})$. However, since we utilize the model-independent GA method the evaluation of this quantity is a rather difficult task.

In order to avoid the dependence on the early time physics of recombination, we then minimize numerically the χ^2 over the quantity $r_s h$ and we construct a new χ^2, which is independent of $r_s h$, namely:

$$\chi^2_{\text{BAO}}(\text{GA}) = \min_{r_s h} [\chi^2_{\text{BAO}}(r_s h, \text{GA})]. \tag{34}$$

We find that this is quite fast and does not affect the efficiency of the GA code.

Note however, that the minimization procedure used here is effectively using the frequentist profile likelihood, but an arguably more appropriate (albeit more computationally costly approach), would be to marginalize over this parameter. As the latter would slow down the code significantly, we prefer to use the minimization approach.

C. Cosmic chronometers

The CC data correspond to another quite useful probe, since they can directly constrain the Hubble rate $H(z)$ at different redshifts, optimizing the differential age method. In particular the CC measurements are based on the raw definition of the Hubble parameter as

$$H(z) = -\frac{1}{1+z}\frac{\mathrm{d}z}{\mathrm{d}t}, \qquad (35)$$

avoiding as a result any complex integrations that may appear in other probes such the SNIa or BAO probes. However, it is important to note that the calculation of the ratio dz/dt that appears in the definition (35) is a rather complex procedure.

In the present work we use the compilation of $36H(z)$ data discussed in Ref. [96] and include three additions as it is illustrated in the Table I of the Appendix. Also, our dataset is similar to that of Ref. [97], but with the addition of a few more points. Also, here we do not include the full covariance matrix, for which the main effect from that compilation is to just increase the errors in the points and further complicate our analysis.

In any case, using the data of Table I, the standard maximum likelihood method can also be applied here, by constructing a χ^2 function. So, following the method proposed in Ref. [41], we minimize over the H_0 parameter (we consider it as a nuisance parameter as we did with M_B in the SNIa case) and find the corresponding χ^2 to be

$$\chi^2_{\mathrm{CC}} = A - \frac{B^2}{\Gamma}, \qquad (36)$$

where the parameters A, B, and Γ are given by the following equations

$$A = \sum_i^{N_H} \left(\frac{H_i}{\sigma_{H_i}}\right)^2, \qquad (37)$$

$$B = \sum_i^{N_H} \frac{H_i E_{\mathrm{th}}(z)}{\sigma^2_{H_i}}, \qquad (38)$$

$$\Gamma = \sum_i^{N_H} \left(\frac{E_{\mathrm{th}}(z)}{\sigma_{H_i}}\right)^2, \qquad (39)$$

where $E_{\mathrm{th}}(z)$ is defined as $E_{\mathrm{th}}(z) \equiv H(z)/H_0$ and N_H describes the total number of CC data. By taking the derivative of Eq. (36) with respect to the nuisance parameter H_0 and setting it equal to zero, it is straightforward to show that the minimum is at

$$H_0 = \frac{B}{\Gamma}. \qquad (40)$$

Note that it is in fact easy to show that for any χ^2 that depends quadratically on a parameter, as is the case for the CC data and H_0, marginalization of this parameter with a flat prior and direct minimization over it gives the same result, up to an irrelevant constant. Thus, the approach shown here is equivalent to marginalizing over H_0.

D. Growth rate data

An additional data compilation that we use in the present work corresponds to the growth rate data compilation presented in Table II of the Appendix (a compilation similar to the one used in Ref. [109]). These data are usually known in the literature as redshift space distortion (RSD) data, due to a particular phenomenon that occurs at both large and small scales during the observation.

In a nutshell, due to the peculiar velocities of galaxies an overdense region seems squashed in redshift space at large scales, while at small scales an overdense region is elongated along the line of sight affecting as a result the two-point correlation function leading to an anisotropic power spectrum.

However, on large scales, a part of the observed anisotropy of the power spectrum can also be due to the use of an incorrect fiducial cosmology $H(z)$ that needs to be taken into account when analyzing the growth rate data and corresponds to the so-called Alcock-Paczynski (AP) effect. In the present analysis, we adopt the rough estimate of Ref. [8], however many alternative forms have been discussed in the literature.

In order to incorporate the RSD data into our analysis, we move along the lines of Ref. [41] and perform a marginalization process over the constant parameters that appear in $f\sigma_8$ and just rescale the values of the theoretical prediction. If we write $f\sigma_8 = \frac{\sigma_8}{\delta_m(1)} a\delta'_m(a) = f\sigma_{8,0} fs(a)$ where $fs(a) = a\delta'_m(a)$, then we can marginalize over the (a priori unknown) scaling constants $f\sigma_{8,0} \equiv \frac{\sigma_8}{\delta_m(1)}$. At this point we also include a correction due to the AP effect, as discussed in Refs. [7,8]. In particular, the correction can be approximated as [120]

$$f\sigma_8(a) \simeq \frac{H(a)D_A(a)}{\tilde{H}(a)\widetilde{D_A}(a)} \tilde{f\sigma}_8(a), \qquad (41)$$

where $H(a)$, $D_A(a)$ are the Hubble parameter and angular diameter distance of the model at hand, $\tilde{H}(a)$, $\tilde{D}_A(a)$ are the Hubble parameter and angular diameter distance of the fiducial cosmology used in the derivation of the data, usually the flat cosmological constant model with some reference $\Omega_{\mathrm{m},0}$ value (given for our data in Table II) and $\tilde{f\sigma}_8(a)$ is the reference value of the data, to be corrected.

While RSD surveys account for the AP effect of each individual point, we still need to correct for it if the cosmology assumed by the survey is quite far from the one where we evaluate $f\sigma_8$. To make matters even more

complicated, as seen in Table II, the fiducial cosmologies assumed by the RSD surveys are quite different from each other. Thus, as it has been noted in the literature, see Ref. [120] and references therein, applying an AP correction again can account (to some extent) for any discrepancies caused by the various $\Omega_{m,0}$ values.

Then we have that $fs(a) = a\delta'(a)\frac{H(a)D_A(a)}{\tilde{H}(a)\widetilde{D_A}(a)}$ and to do the marginalization, we write the χ^2 as

$$\chi^2 = [f\sigma_{8,i} - f\sigma_{8,0}fs(z_i)]C_{ij}^{-1}[f\sigma_{8,j} - f\sigma_{8,0}fs(z_j)],$$
$$= A - f\sigma_{8,0}B + f\sigma_{8,0}^2\Gamma, \tag{42}$$

where we have expanded the sum and defined the coefficients A, B, Γ as

$$A = f\sigma_{8,i}C_{ij}^{-1}f\sigma_{8,j}, \tag{43}$$

$$B = f\sigma_{8,i}C_{ij}^{-1}fs(z_j) + fs(z_i)C_{ij}^{-1}f\sigma_{8,j}, \tag{44}$$

$$\Gamma = fs(z_i)C_{ij}^{-1}fs(z_j). \tag{45}$$

Then the value of $f\sigma_{8,0}$ at the minimum and the corresponding marginalized χ^2 are, respectively,

$$f\sigma_{8,0\,\min} = \frac{B}{2\Gamma}, \tag{46}$$

$$\tilde{\chi}^2 = A - \frac{B^2}{4\Gamma}, \tag{47}$$

which is what we use in our analysis. As mentioned earlier, as the parameter $f\sigma_{8,0}$ appears quadratically in the χ^2, then minimizing over it is in fact equivalent to marginalizing over it.

Finally, even though the RSD and BAO points come in principle from the same or related datasets, we cannot take into account these possible correlations as in general, we have no access to the necessary covariance matrices. Even in the cases of the eBOSS data where we have the covariance between the RSD and BAO point, it is difficult to incorporate the covariance matrix into the analysis as we perform the GA fitting separately. Thus, to keep the analysis simple we will assume we can just add the RSD and BAO χ^2 terms together.

E. E_g data

Last but not least, we have used an updated compilation of seven uncorrelated E_g data points that are presented in the Table III of the Appendix. These consist of five weak gravitational lensing datapoints from the Kilo-Degree Survey (KiDS-1000), juxtaposed with overlapping data from the BOSS and 2dFLenS galaxy spectroscopic redshift surveys [121] and two datapoints from the VIMOS Public Extragalactic Redshift Survey (VIPERS) [122].

The E_g statistic as expressed via Eq. (15) is galaxy bias independent at linear order, since it was created by definition as a probe of the ratio of the Newtonian potentials (Φ, Ψ) of the perturbed FLRW metric given by Eq. (2). In order to fit the E_g data we can construct and minimize a χ^2 formula of the form

$$\chi^2_{E_g} = \sum_i \left(\frac{2E_g^i - P_2(a^i)}{2\sigma_{E_g,i}}\right)^2. \tag{48}$$

Furthermore, in our analysis we use the GA pipeline to directly fit the function $P_2(a)$ given analytically from Eq. (9).

In general, the E_g data have been known to be plagued by scale and bias-dependent lensing contributions, which in effect increase the systematic uncertainties [123]. While this can be ameliorated to a certain extend by adding the correlations of shear and galaxy clustering, it is unclear whether the current data have these corrections [124].

One possibility to at least take into account this extra uncertainty is to introduce an intrinsic systematic error σ_{stat} that has to be determined from the data such that the χ^2 per degree of freedom (d.o.f.) is order unity, i.e., $\chi^2/\text{d.o.f.} \sim 1$, as was done in the past for some SNIa compilations (see Ref. [125]), but this is a rather questionable statistical practice as it assumes *a priori* the validity of the model and makes model comparison impossible.

Thus, in what follows we will interpret any results stemming from the currently available E_g data with caution, even if they are very promising and may already weakly hint towards some new physics, as it is uncertain if the possible systematics can explain the observed deviations from GR [41].

F. The mock data

Here we briefly discuss the mock data we used in our analysis. In particular, we consider two different fiducial cosmologies, one based on the ΛCDM model and another one based on a model with an evolving Newton's constant $\mu(k,a)$ and lensing parameter $\Sigma(k,a)$, as defined via the Poisson equations Eqs. (3) and (4). We call the latter model the μCDM. Specifically, we use the parametrizations:

$$\mu(k,a) = 1 + g_a(1-a)^{m_1} - g_a(1-a)^{2m_1}, \tag{49}$$

$$\Sigma(k,a) = 2 + \sigma_a(1-a)^{m_2} - \sigma_a(1-a)^{2m_2}, \tag{50}$$

where g_a, σ_a are some parameters ($g_a = \sigma_a = 0$ in the ΛCDM model), and we set $m_1 = m_2 = 2$, inspired by Ref. [7] such that the models pass the Solar System tests. The parameters for the μCDM are chosen to be $g_a = -0.627$ and $\sigma_a = -3.562$, which are the best-fit values

found in Ref. [10] (see the "Datasets $f\sigma_8 + E_g$ corr." combination in their Table IV).

For the rest of the cosmological parameters (which are common to both fiducial cosmologies) we assume $\Omega_{m,0} = 0.3$, $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 0.7$, $\sigma_{8,0} = 0.8$ and for the dark energy parameters $(w_0, w_a) = (-1, 0)$, i.e., we assume a ΛCDM background as we are interested in the effects of the cosmological perturbations.

Having fixed the cosmological parameters for the two models, we then make the corresponding two sets of data for the Hubble parameter $H(z)$, the growth rate $f\sigma_8(z)$ and the $E_g(z)$ parameter for each model in 20 redshift bins in the range $z \in [0, 2.0]$ with a fixed bin size of $dz = 0.1$. We also add Gaussian noise and errors to the 1% level, which is within the expectations of forthcoming next-generation surveys, such as Euclid [126].

In fact, Euclid will obtain precise measurements of these quantities, albeit its redshift range is much more limited, e.g., for the spectroscopic survey the range is $z \in [0.9, 1.8]$ and needs to be complemented by other surveys (e.g., DESI) in lower redshifts [58–60]. Thus, we consider our methodology as a proof of concept approach and the forecasts as an optimistic case to gauge how well the method works overall.

With these two mock data sets at hand, we then rerun the GA pipelines as in the previous section and we again reconstruct the parameters $\mu(k, a)$ and $\eta(k, a)$, aiming to forecast at which level our ML approach is able to detect deviations from GR. One key difference from the analysis of the real data in the previous section is though that in this case we do not require any SNIa data or any corrections for the AP effect in the growth data, as all the points will come from the same source, unlike the compilation of the currently available points used here, thus allowing us to perform a much cleaner analysis.

We discuss the results of our analysis in the next section in detail.

V. RESULTS

Using the methodology described earlier, in order to directly reconstruct $\mu(z)$ and $\eta(z)$ we aim to produce analytic fits for the functions $H(z)$, $D_L(z)$, $D_A(z)$, $f\sigma_8(z)$, and $E_g(z)$, via updated compilations of the CC, SNIa, BAO, RSD, and E_g data for the real data. On the other hand, fit for the functions $H(z)$, $f\sigma_8(z)$, and $E_g(z)$ in the case of the mock data.

Before running the GA analysis we choose a specific grammar and we also impose a few physically motivated hypotheses in order to ensure that the obtained functions are well-behaved, e.g., smooth, continuous, and singularity free as expected by physical quantities. In particular, we impose the following priors:

(i) The luminosity distance $D_L(z)$ in the low redshift regime should be approximated by the Hubble law

as $D_L(z \to 0) \simeq cz/H_0$. The same behavior is also expected for the angular diameter distance $D_A(z)$ in the low redshift regime, since the two quantities in the context of FLRW metric are connected via the standard distance duality relation

$$D_L(z) = (1 + z)^2 D_A(z). \tag{51}$$

(ii) The present-day value of the Hubble parameter should be equal to the Hubble constant H_0, which is evaluated via the CC data.

(iii) In the high redshift regime, i.e., deep in the matter domination era, the growth factor $\delta_m(a)$ should evolve as $\delta_m(a) \simeq a$, as expected from LSS observations.

The particular prior for the distances is implemented by demanding that the luminosity distance behaves as $D_L(z) = cH_0^{-1}z[1 + z \cdot \text{GA}(z)]$, where $\text{GA}(z)$ is the function predicted by the GA. This ensures that the prior is actually enforced at all times for well-behaved functions (as are the ones created by our implementation of the GA). Also, in order to maximize the freedom in the reconstructions in this work we follow Ref. [41] and perform a separate analysis by using two separate GA functions for $D_L(z)$ and $H(z)$.

However, for completeness we also compared the expressions for the luminosity distance: one determined directly by the GA and the other by integrating the Hubble parameter. Doing so we find that in the redshift range $z \in [0, 2]$ the agreement between the two approaches is on average $\sim 1.2\%$ and always below 2%. Thus, there is reasonable agreement between the two function, even if we do not enforce Eq. (19).

On the other hand, in the high redshift regime, i.e., deep in the matter domination era, while the growth factor $\delta_m(a)$ should evolve as $\delta_m(a) \simeq a$, as expected from LSS observations, these high redshifts ($z \in [10^2, 10^4]$) are not yet probed by LSS observations and are only indirectly constrained by the CMB, so any deviations from the expected behavior are only speculative at the moment. Still, in order to probe for any deviations, we extend the GA grammar used, by considering also polynomials of fractional powers of the scale factor a.

Finally, the initial functions are then subjected to the crossover and mutation operations and after this process is repeated several times (until we reach a maximum number of generations to ensure convergence), the final expressions for the best-fit functions are obtained.

A. Currently available data

Running the GA pipelines for the currently available data, the following best-fit function forms are obtained:

$$\left(\frac{H(z)}{H_0}\right)^2 = 1 + z[0.262(0.153z+1)^3 - 0.618z - 1.281]^2, \tag{52}$$

$$D_L(z) = \frac{c}{H_0} z[1 + z(0.872 - 0.129z)^2], \tag{53}$$

$$f\sigma_8(a) = f_0\{a - a^4[1.680(0.073^{0.073a}a^{0.073a})^{1.720} + 0.003a^{5.680} - 0.412a^{1.533}]^2\}, \tag{54}$$

$$P_2(a) = (-0.160 \cdot 0.180^{0.361a}a^{0.361a} + 4.758a^9 - 2.869a^3)^2, \tag{55}$$

where $f_0 = 1.109 \pm 0.272$ and $H_0 = (67.17 \pm 12.22)$ km s^{-1} Mpc^{-1}, which is compatible with the Planck 2018 measurements. The redshift evolution of the growth rate $f\sigma_8(z)$ and of the quantity $P_2(z)$ as reconstructed by the GA using the RSD and E_g data, is illustrated in Fig. 2 (blue and brown lines), while the evolution of the Hubble rate $H(z)$, of the luminosity distance $D_L(z)$ and of the angular diameter distance $D_A(z)$ as reconstructed by the GA using the CC, the BAO, and the SNIa data, is illustrated in Figs. 3 and 4 accordingly (red, green, and magenta lines). In each plot we also include the best fit of the ΛCDM scenario (dashed black lines) as well as the 1σ errors regions (shaded areas) utilizing the path integral approach.

Clearly, the ΛCDM case is consistent with the results of the GA, i.e., the black dashed line is allocated well within $\sim 1\sigma$ in almost all of the cases except the results produced from the E_g data, where the $P_2(z)$ function as derived by the GA pipeline presents a large deviation from the one predicted for the ΛCDM case. However, it is important to note that even the ΛCDM scenario does not fit well the E_g data compilation, the fact that the χ^2 per d.o.f. is $\simeq 1.8$, which is an indication that the E_g results should be
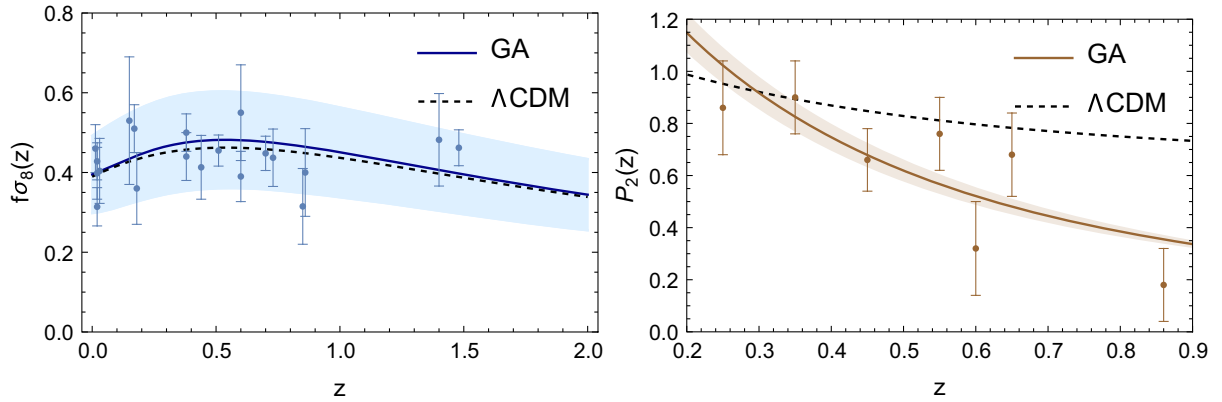

FIG. 2. The best-fit $f\sigma_8(z)$ (left) and $P_2(z)$ (right) functions (blue and brown lines), with their respective data points, as predicted by the GA pipelines using growth and E_g data accordingly, along with their 1σ error bands (light blue and light brown areas). We see that while on one hand, the predicted $f\sigma_8(z)$ function is consistent with the one produced assuming ΛCDM (dashed black line) the same cannot be said for the predicted $P_2(z)$ function, which is affected by the low quality of the data.
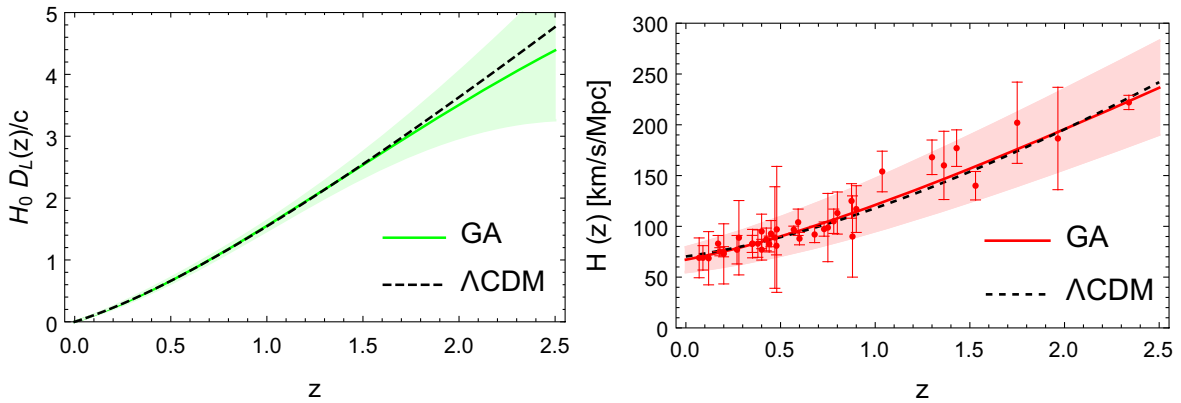

FIG. 3. The best-fit for the luminosity distance $D_L(z)$ (left) and Hubble rate $H(z)$ (right) functions (green and red lines) as predicted by the GA using CC and SNIa data accordingly, with the CC data points, along with their 1σ error bands (light green and light red areas). The dashed black line corresponds to the ΛCDM best fit.
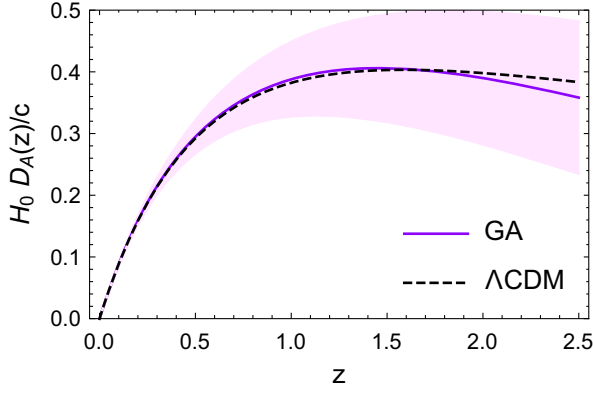
FIG. 4.   The best-fit angular diameter distance $D_A(z)$ function (magenta line) as predicted by the GA using a compilation of BAO data, along with its 1σ error band (light magenta area). The dashed black line corresponds to the ΛCDM best fit.

interpreted with care. Interestingly, a similar deviation from the ΛCDM case was also observed for $P_2(z)$ in Ref. [41].

Using the aforementioned GA reconstructions we finally calculate the evolution of the parameter $\Omega_{\mathrm{m},0}\mu(z)$ using the theoretical expressions presented in Sec. II. The final result is shown in the left panel of Fig. 5 where we show the GA reconstruction in a solid black line along with the 1σ errors as the shaded region. Overall we find that while there is good agreement with the expectation of a constant value, but we find that the errors are quite large due to the lower quality of the currently available data compared to the ones from forthcoming surveys in the near future, thus not allowing us to draw any strong conclusions on any deviations of $\Omega_{\mathrm{m},0}\mu(z)$ from a constant value.

We also perform the GA reconstruction of the anisotropic stress η, which is shown in the right panel of Fig. 5 in the redshift range $z \in [0.2, 0.9]$ where we have the E_g data, but we find that it deviates strongly from unity. The reason for this is that the E_g data are most likely plagued with systematics (as discussed in detail in Sec. IV E), which most likely drive the deviation seen. Of course, the possibility of new physics also cannot be excluded as the systematics in the data are unlikely to account for the whole deviation from unity observed in η. A similar result was found for η in Ref. [41] using an earlier compilation of the data.
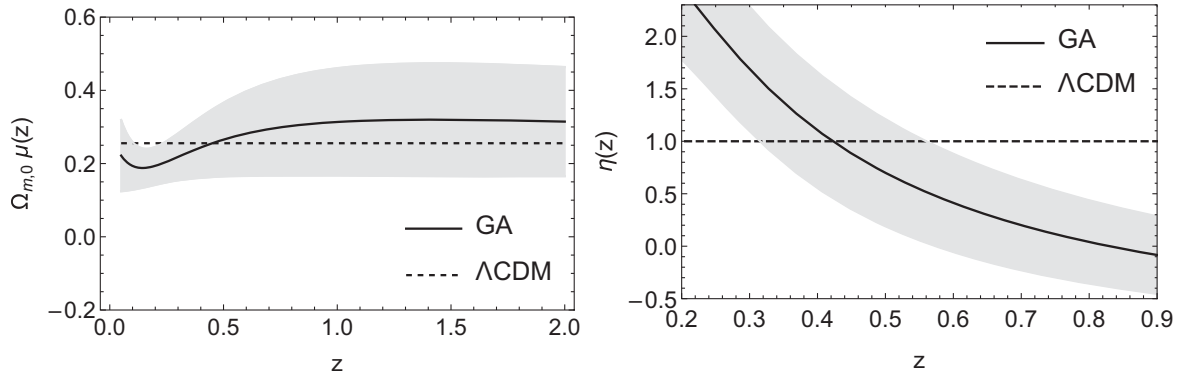

FIG. 5.   The reconstructions of $\Omega_{\mathrm{m},0}\mu(z)$ (left) and $\eta(z)$ (right) given by Eqs. (14) and (12), respectively, where the relevant functions were produced by the GA pipelines. The solid black lines correspond to the GA reconstruction, while the 1σ error band is the light gray area. Note that we only show the reconstruction of η in the redshift range $z \in [0.2, 0.9]$ where we have the E_g data.
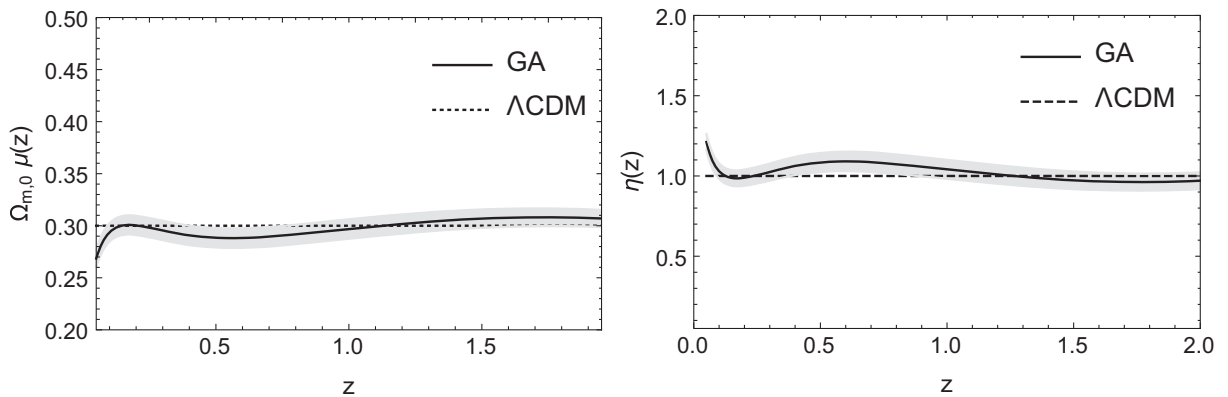

FIG. 6.   The GA reconstruction of $\Omega_{\mathrm{m},0}\mu(z)$ (left) and $\eta(z)$ (right) for the ΛCDM mock. The gray-shaded regions correspond to the 1σ confidence level, while the dashed red line corresponds to the fiducial model $\mu(z) = \eta(z) = 1$, i.e., GR and the ΛCDM model. In both cases the GA is able to constrain the correct underlying fiducial model to within a few percent.
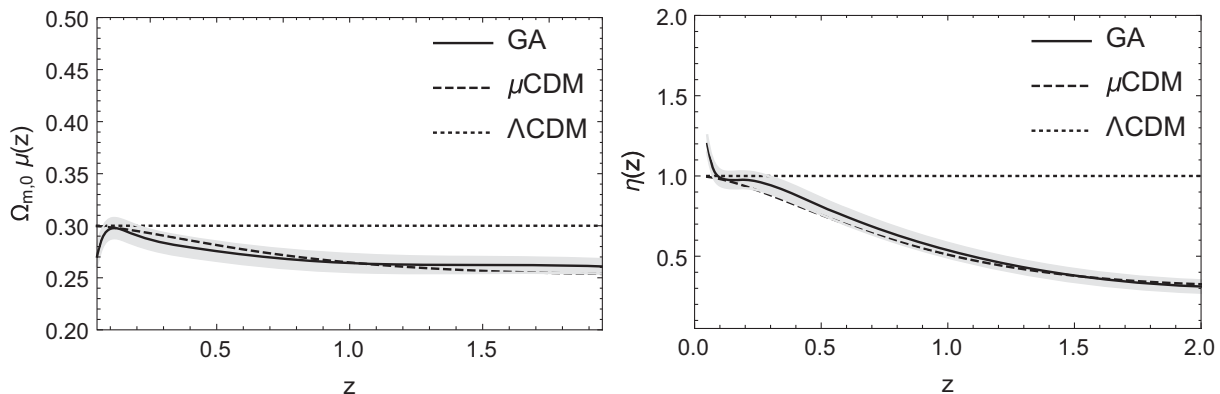
FIG. 7. The GA reconstruction of $\Omega_{\mathrm{m},0}\mu(z)$ (left) and $\eta(z)$ (right) for the μCDM mock. The gray-shaded regions correspond to the 1σ confidence level, while the dashed red line corresponds to the fiducial model described by Eqs. (49) and (50) and the parameters $g_a = -0.627$ and $\sigma_a = -3.562$ for $m_1 = m_2 = 2$.

B. Mock data

Here we repeat the GA reconstruction analysis, however this time using the mock data described in Sec. IV F. In particular we run the whole GA pipeline on the ΛCDM and μCDM mocks and from there we reconstruct the quantity $\Omega_{\mathrm{m},0}\mu(z)$ which is related to the effective Newton's constant and the anisotropic stress parameter $\eta(z)$ using Eqs. (16) and (17), respectively.

In Fig. 6 we show the results of these reconstructions for the ΛCDM mock. In particular, we show $\Omega_{\mathrm{m},0}\mu(z)$ in the left panel and $\eta(z)$ on the right panel of Fig. 6 respectively and as can be seen, in both cases the GA is able to recover the correct underlying cosmology within the errors, given by the gray-shaded regions at the 1σ confidence level.

On the other hand, in Fig. 7 we show the GA reconstructions for the μCDM mock, again $\Omega_{\mathrm{m},0}\mu(z)$ in the left panel and $\eta(z)$ in the right panel, respectively. Here we find that even though the fiducial cosmology is quite extreme, the GA is able to reconstruct both functions to within a few percent and provide a detection, quantified as a deviation from a constant value, at several σs.

VI. CONCLUSIONS

In this work, we used a particular ML approach, called the GA, in order to perform nonparametric reconstructions of two key quantities that parametrize deviations from GR, namely the effective Newton's constant G_{eff} and the anisotropic parameter η [127,128].

To perform the reconstructions we used both the currently available data, coming from the BAO, the CC, the growth of matter perturbations, and the so-called E_g statistics compilations but also synthetic data assuming a next-generation survey in order to forecast, in an optimist scenario, how well we will be able to constrain deviations

from the null hypothesis, either the ΛCDM model or a model with an evolving Newton's constant, dubbed μCDM.

In the case of the currently available data we find, as expected, that the GA reconstructions are affected by the lower quality of the currently available data. While $\Omega_{\mathrm{m},0}\mu(z)$ seems to be in good agreement with having a constant value at all redshifts within the errors, on the other hand the anisotropic stress η is more difficult to interpret as the fit is plagued by the systematics of the E_g data, in agreement with previous works. While it might have been interesting to repeat the same analysis without the E_g data, it is in fact not possible to do that as $\Omega_{\mathrm{m},0}\mu(z)$ does not contain the E_g function but $\eta(z)$ requires it, thus we cannot do a separate analysis.

The situation is more clear when using the fiducial data based on the two different mocks. Here we find that in both cases the GA reconstructions are in good agreement with the fiducial models, either the ΛCDM for which $\mu = \eta = 1$ or the μCDM model, which had a more complicated evolution.

In particular, we see in Figs. 6 and 7 that the GA reconstructions are well within the 1σ errors of the true fiducial model and the GA is able to constrain the correct underlying fiducial model to within a few percent. This is possible for both mocks, and in the case of the μCDM mock, which was based on a rather extreme cosmology, can provide a strong detection of several σs.

Overall, we find that this reconstruction method is a very useful tool in performing model-independent reconstructions of the two key quantities $\Omega_{\mathrm{m},0}\mu(z)$ and $\eta(z)$ that can be used to parametrize most modified gravity models and for which any deviations from unity would be smoking guns for new physics if the systematics are under control.

Numerical analysis files.—The *Mathematica* codes used by the authors in the analysis of the paper will be made publicly available upon publication at https://github.com/snesseris/.

APPENDIX: DATA COMPILATIONS

We present the updated compilations of cosmological data that were used in the analysis.

TABLE I.   A compilation of 39 cosmic chronometer data dating from 2009 to 2022, in units of $\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

Index	z	$H(z) \pm \sigma_H$	References	Date
1	0.09	69 ± 12	[98]	July 2009
2	0.17	83 ± 8	[98]	July 2009
3	0.27	77 ± 14	[98]	July 2009
4	0.40	95 ± 17	[98]	July 2009
5	0.48	97 ± 62	[98]	July 2009
6	0.88	90 ± 40	[98]	July 2009
7	0.90	117 ± 23	[98]	July 2009
8	1.30	168 ± 17	[98]	July 2009
9	1.43	177 ± 18	[98]	July 2009
10	1.53	140 ± 14	[98]	July 2009
11	1.75	202 ± 40	[98]	July 2009
12	0.44	82.6 ± 7.8	[87]	June 2012
13	0.60	87.9 ± 6.1	[87]	June 2012
14	0.73	97.3 ± 7.0	[87]	June 2012
15	0.179	75 ± 4	[99]	February 2013
16	0.199	75.0 ± 5	[99]	February 2013
17	0.352	83.0 ± 14	[99]	February 2013
18	0.593	104.0 ± 13	[99]	February 2013
19	0.68	92.0 ± 8	[99]	February 2013
20	0.781	105.0 ± 12	[99]	February 2013
21	0.875	125.0 ± 17	[99]	February 2013
22	1.037	154.0 ± 20	[99]	February 2013
23	0.35	82.7 ± 8.4	[100]	August 2013
24	0.07	69.0 ± 19.6	[101]	May 2014
25	0.12	68.6 ± 26.2	[101]	May 2014
26	0.20	72.9 ± 29.6	[101]	May 2014
27	0.28	88.8 ± 36.6	[101]	May 2014
28	0.57	96.8 ± 3.4	[102]	June 2014
29	2.34	222.0 ± 7.0	[103]	December 2014
30	1.363	160.0 ± 33.6	[104]	March 2015
31	1.965	186.5 ± 50.4	[104]	March 2015
32	0.3802	83.0 ± 13.5	[105]	May 2016
33	0.4004	77.0 ± 10.2	[105]	May 2016
34	0.4247	87.1 ± 11.2	[105]	May 2016
35	0.4497	92.8 ± 12.9	[105]	May 2016
36	0.4783	80.9 ± 9.0	[105]	May 2016
37	0.47	89 ± 50	[106]	February 2017
38	0.75	98.8 ± 33.6	[107]	October 2021
39	0.80	113.1 ± 20.73	[108]	May 2022

TABLE II.   An updated compilation of RSD data. The correlations pertaining to the eBOSS data at $z = 0.38$ and $z = 0.51$ are explained here.

Index	Survey	z	$f\sigma_8 \pm \sigma_{f\sigma_8}$	$\Omega_{m,0}$	References	Date
1	2dFGRS	0.17	0.510 ± 0.060	0.3	[110]	July 2008
2	2MASS	0.02	0.314 ± 0.048	0.266	[111,112]	January 2011
3	SNIa + IRAS	0.02	0.398 ± 0.065	0.3	[112,113]	November 2011
4	WiggleZ	0.44	0.413 ± 0.080	0.27	[87]	April 2012
5	WiggleZ	0.60	0.390 ± 0.063	0.27	[87]	April 2012
6	WiggleZ	0.73	0.437 ± 0.072	0.27	[87]	April 2012
7	GAMA	0.18	0.36 ± 0.09	0.27	[114]	September 2013
8	GAMA	0.38	0.44 ± 0.06	0.27	[114]	September 2013
9	FastSound	1.4	0.482 ± 0.116	0.27	[115]	March 2016
10	6dFGS + SNIa	0.02	$0.428^{+0.048}_{-0.045}$	0.3	[116]	November 2016
11	VIPERS PDR2	0.6	0.55 ± 0.12	0.3	[117]	December 2016
12	VIPERS PDR2	0.86	0.40 ± 0.11	0.3	[117]	December 2016
13	2MTF + 6dFGSv	0.03	$0.404^{+0.082}_{-0.081}$	0.312	[118]	June 2019
14	ALFALFA	0.013	0.46 ± 0.06	0.315	[119]	May 2021
15	eBOSS	0.15	0.53 ± 0.16	0.31	[91]	July 2020
16	eBOSS	0.38	0.500 ± 0.047	0.31	[91]	July 2020
17	eBOSS	0.51	0.455 ± 0.039	0.31	[91]	July 2020
18	eBOSS	0.70	0.448 ± 0.043	0.31	[91]	July 2020
19	eBOSS	0.85	0.315 ± 0.095	0.31	[91]	July 2020
20	eBOSS	1.48	0.462 ± 0.045	0.31	[91]	July 2020

TABLE III.   The updated and uncorrelated compilation of E_g data from 2016 to 2020.

Index	Survey	z	$E_g \pm \sigma_{E_g}$	References	Date
1	VIPERS	0.6	0.16 ± 0.09	[122]	December 2016
2	VIPERS	0.86	0.09 ± 0.07	[122]	December 2016
3	KiDS-1000 + BOSS DR12 + 2dFLenS	0.25	0.43 ± 0.09	[121]	May 2020
4	KiDS-1000 + BOSS DR12 + 2dFLenS	0.35	0.45 ± 0.07	[121]	May 2020
5	KiDS-1000 + BOSS DR12 + 2dFLenS	0.45	0.33 ± 0.06	[121]	May 2020
6	KiDS-1000 + BOSS DR12 + 2dFLenS	0.55	0.38 ± 0.07	[121]	May 2020
7	KiDS-1000 + BOSS DR12 + 2dFLenS	0.65	0.34 ± 0.08	[121]	May 2020

[1] L. Perivolaropoulos and F. Skara, New Astron. Rev. **95,** 101659 (2022).

[2] E. N. Saridakis *et al.* (CANTATA Collaboration), arXiv: 2105.12582.

[3] E. Abdalla *et al.*, J. High Energy Astrophys. **34,** 49 (2022).

[4] S. Joudaki *et al.*, Mon. Not. R. Astron. Soc. **474,** 4894 (2018).

[5] T. M. C. Abbott *et al.* (DES Collaboration), Phys. Rev. D **98,** 043526 (2018).

[6] S. Basilakos and S. Nesseris, Phys. Rev. D **96,** 063517 (2017).

[7] S. Nesseris, G. Pantazis, and L. Perivolaropoulos, Phys. Rev. D **96,** 023542 (2017).

[8] L. Kazantzidis and L. Perivolaropoulos, Phys. Rev. D **97,** 103503 (2018).

[9] L. Perivolaropoulos and L. Kazantzidis, Int. J. Mod. Phys. D **28,** 1942001 (2019).

[10] F. Skara and L. Perivolaropoulos, Phys. Rev. D **101,** 063521 (2020).

[11] G. Alestas and L. Perivolaropoulos, Mon. Not. R. Astron. Soc. **504,** 3956 (2021).

[12] A. Rassat, J. L. Starck, P. Paykari, F. Sureau, and J. Bobin, J. Cosmol. Astropart. Phys. 08 (2014) 006.

[13] D. J. Schwarz, C. J. Copi, D. Huterer, and G. D. Starkman, Classical Quantum Gravity **33,** 184001 (2016).

[14] Y. Akrami *et al.* (Planck Collaboration), Astron. Astrophys. **641,** A7 (2020).

[15] L. Perivolaropoulos, Galaxies **2,** 22 (2014).

[16] J. Muir, S. Adhikari, and D. Huterer, Phys. Rev. D **98,** 023521 (2018).

[17] A. E. Bayer and U. Seljak, J. Cosmol. Astropart. Phys. 10 (2020) 009.

[18] R. H. Cyburt, B. D. Fields, and K. A. Olive, Phys. Lett. B **567,** 227 (2003).

[19] R. H. Cyburt, B. D. Fields, and K. A. Olive, J. Cosmol. Astropart. Phys. 11 (2008) 012.

[20] V. Poulin and P. D. Serpico, Phys. Rev. Lett. **114,** 091101 (2015).

[21] K. Mori and M. Kusakabe, Phys. Rev. D **99,** 083013 (2019).

[22] S. Hayakawa *et al.*, J. Phys. Soc. Jpn. Conf. Proc. **31,** 011036 (2020).

[23] S. Ishikawa *et al.*, J. Phys. Soc. Jpn. Conf. Proc. **31,** 011037 (2020).

[24] M. Ishak, Living Rev. Relativity **22,** 1 (2019).

[25] L. Kazantzidis and L. Perivolaropoulos, σ_8 tension. Is gravity getting weaker at low z? Observational evidence and theoretical implications, in *Modified Gravity and Cosmology* (Springer Nature, Switzerland, 2021).

[26] E. Di Valentino, O. Mena, S. Pan, L. Visinelli, W. Yang, A. Melchiorri, D. F. Mota, A. G. Riess, and J. Silk, Classical Quantum Gravity **38,** 153001 (2021).

[27] N. Schöneberg, G. Franco Abellán, A. Pérez Sánchez, S. J. Witte, V. Poulin, and J. Lesgourgues, Phys. Rep. **984,** 1 (2022).

[28] G. Alestas, L. Kazantzidis, and L. Perivolaropoulos, Phys. Rev. D **101,** 123516 (2020).

[29] P. Shah, P. Lemos, and O. Lahav, Astron. Astrophys. Rev. **29,** 9 (2021).

[30] N. Aghanim *et al.* (Planck Collaboration), Astron. Astrophys. **641,** A6 (2020); **652,** C4(E) (2021).

[31] A. G. Riess *et al.*, Astrophys. J. Lett. **934,** L7 (2022).

[32] G. Alestas, L. Kazantzidis, and L. Perivolaropoulos, Phys. Rev. D **103,** 083517 (2021).

[33] V. Marra and L. Perivolaropoulos, Phys. Rev. D **104,** L021303 (2021).

[34] G. Alestas, I. Antoniou, and L. Perivolaropoulos, Universe **7,** 366 (2021).

[35] L. Perivolaropoulos and F. Skara, Phys. Rev. D **104,** 123511 (2021).

[36] G. Alestas, L. Perivolaropoulos, and K. Tanidis, Phys. Rev. D **106,** 023526 (2022).

[37] L. Perivolaropoulos and F. Skara, Phys. Rev. D **106,** 043528 (2022).

[38] S. D. Odintsov and V. K. Oikonomou, Europhys. Lett. **137,** 39001 (2022).

[39] L. Amendola, P. S. Corasaniti, and F. Occhionero, arXiv:astro-ph/9907222.

[40] E. Gaztanaga, E. Garcia-Berro, J. Isern, E. Bravo, and I. Dominguez, Phys. Rev. D **65,** 023506 (2001).

[41] R. Arjona and S. Nesseris, J. Cosmol. Astropart. Phys. 11 (2020) 042.

[42] P. Zhang, M. Liguori, R. Bean, and S. Dodelson, Phys. Rev. Lett. **99,** 141302 (2007).

[43] L. Amendola, M. Kunz, M. Motta, I. D. Saltas, and I. Sawicki, Phys. Rev. D **87,** 023501 (2013).

[44] A. M. Pinho, S. Casas, and L. Amendola, J. Cosmol. Astropart. Phys. 11 (2018) 027.

[45] S. Tsujikawa, Phys. Rev. D **76,** 023514 (2007).

[46] C.-P. Ma and E. Bertschinger, Astrophys. J. **455,** 7 (1995).

[47] I. D. Saltas and M. Kunz, Phys. Rev. D **83,** 064042 (2011).

[48] L. Amendola, S. Fogli, A. Guarnizo, M. Kunz, and A. Vollmer, Phys. Rev. D **89,** 063538 (2014).

[49] I. Sawicki, I. D. Saltas, L. Amendola, and M. Kunz, J. Cosmol. Astropart. Phys. 01 (2013) 004.

[50] S. Nesseris, C. Blake, T. Davis, and D. Parkinson, J. Cosmol. Astropart. Phys. 07 (2011) 037.

[51] C. Bogdanos and S. Nesseris, J. Cosmol. Astropart. Phys. 05 (2009) 006.

[52] S. Nesseris and A. Shafieloo, Mon. Not. R. Astron. Soc. **408,** 1879 (2010).

[53] S. Nesseris and J. Garcia-Bellido, J. Cosmol. Astropart. Phys. 11 (2012) 033.

[54] S. Nesseris and J. García-Bellido, Phys. Rev. D **88,** 063521 (2013).

[55] D. Sapone, E. Majerotto, and S. Nesseris, Phys. Rev. D **90,** 023012 (2014).

[56] R. Arjona, J. Cosmol. Astropart. Phys. 08 (2020) 009.

[57] R. Arjona and S. Nesseris, Phys. Rev. D **101,** 123525 (2020).

[58] M. Martinelli *et al.* (EUCLID Collaboration), Astron. Astrophys. **644,** A80 (2020).

[59] M. Martinelli *et al.* (Euclid Collaboration), Astron. Astrophys. **654,** A148 (2021).

[60] S. Nesseris *et al.* (Euclid Collaboration), Astron. Astrophys. **660,** A67 (2022).

[61] S. Abel, D. G. Cerdeño, and S. Robles, arXiv:1805.03615.

[62] B. C. Allanach, D. Grellscheid, and F. Quevedo, J. High Energy Phys. 07 (2004) 069.

[63] Y. Akrami, P. Scott, J. Edsjo, J. Conrad, and L. Bergstrom, J. High Energy Phys. 04 (2010) 057.

[64] M. Wahde and K. Donner, Astron. Astrophys. **379,** 115 (2001).

[65] V. Rajpaul, in *56th Annual Conference of the South African Institute of Physics* (2012), pp. 519–524, arXiv:1202.1643.

[66] M. Ho, M. M. Rau, M. Ntampaka, A. Farahi, H. Trac, and B. Poczos, Astrophys. J. **887,** 25 (2019).

[67] M. Affenzeller, S. Wagner, S. Winkler, and A. Beham, *Genetic Algorithms and Genetic Programming: Modern Concepts and Practical Applications* (Chapman and Hall/CRC, London, 2009).

[68] S. Sivanandam and S. Deepa, in *Introduction to Genetic Algorithms* (Springer, New York, 2008), pp. 15–37.

[69] S.-M. Udrescu and M. Tegmark, Sci. Adv. **6,** eaay2631 (2020).

[70] Y. Setyawati, M. Pürrer, and F. Ohme, Classical Quantum Gravity **37,** 075012 (2020).

[71] H. Vaddireddy, A. Rasheed, A. E. Staples, and O. San, Phys. Fluids **32,** 015113 (2020).

[72] K. Liao, A. Shafieloo, R. E. Keeley, and E. V. Linder, Astrophys. J. Lett. **886,** L23 (2019).

[73] E. Belgacem, S. Foffa, M. Maggiore, and T. Yang, Phys. Rev. D **101,** 063505 (2020).

[74] Y. Li, R. S. Rainer, W. Cheng, and Y. Hao, in *3rd North American Particle Accelerator Conference (NAPAC2019)*

(JACoW Publishing, Geneva, Switzerland, 2019), pp. 912–915.

[75] M. Bernardini, L. Mayer, D. Reed, and R. Feldmann, Mon. Not. R. Astron. Soc. **496,** 5116 (2020).

[76] A. Gómez-Valent and L. Amendola, in *15th Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics, and Relativistic Field Theories* (2019), arXiv:1905.04052.

[77] R. Arjona, H.-N. Lin, S. Nesseris, and L. Tang, Phys. Rev. D **103,** 103513 (2021).

[78] D. M. Scolnic *et al.* (Pan-STARRS1 Collaboration), Astrophys. J. **859,** 101 (2018).

[79] D. Zhao, Y. Zhou, and Z. Chang, Mon. Not. R. Astron. Soc. **486,** 5679 (2019).

[80] L. Kazantzidis and L. Perivolaropoulos, Phys. Rev. D **102,** 023520 (2020).

[81] D. Sapone, S. Nesseris, and C. A. P. Bengaly, Phys. Dark Universe **32,** 100814 (2021).

[82] L. Kazantzidis, H. Koo, S. Nesseris, L. Perivolaropoulos, and A. Shafieloo, Mon. Not. R. Astron. Soc. **501,** 3421 (2021).

[83] M. G. Dainotti, B. De Simone, T. Schiavone, G. Montani, E. Rinaldi, and G. Lambiase, Astrophys. J. **912,** 150 (2021).

[84] A. Conley *et al.* (SNLS Collaboration), Astrophys. J. Suppl. Ser. **192,** 1 (2011).

[85] M. Betoule *et al.* (SDSS Collaboration), Astron. Astrophys. **568,** A22 (2014).

[86] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson, Mon. Not. R. Astron. Soc. **416,** 3017 (2011).

[87] C. Blake *et al.*, Mon. Not. R. Astron. Soc. **425,** 405 (2012).

[88] C. Escamilla-Rivera, Galaxies **4,** 8 (2016).

[89] T. M. C. Abbott *et al.* (DES Collaboration), Phys. Rev. D **105,** 043512 (2022).

[90] H. du Mas des Bourboux *et al.*, Astrophys. J. **901,** 153 (2020).

[91] S. Alam *et al.* (eBOSS Collaboration), Phys. Rev. D **103,** 083533 (2021).

[92] E. Aubourg *et al.*, Phys. Rev. D **92,** 123516 (2015).

[93] D. J. Eisenstein and W. Hu, Astrophys. J. **496,** 605 (1998).

[94] E. Komatsu *et al.* (WMAP Collaboration), Astrophys. J. Suppl. Ser. **180,** 330 (2009).

[95] J. E. Bautista *et al.*, Astron. Astrophys. **603,** A12 (2017).

[96] R. Arjona, W. Cardona, and S. Nesseris, Phys. Rev. D **99,** 043516 (2019).

[97] M. Moresco *et al.*, arXiv:2201.07241.

[98] D. Stern, R. Jimenez, L. Verde, M. Kamionkowski, and S. A. Stanford, J. Cosmol. Astropart. Phys. 02 (2010) 008.

[99] M. Moresco *et al.*, J. Cosmol. Astropart. Phys. 08 (2012) 006.

[100] C.-H. Chuang and Y. Wang, Mon. Not. R. Astron. Soc. **435,** 255 (2013).

[101] C. Zhang, H. Zhang, S. Yuan, T.-J. Zhang, and Y.-C. Sun, Res. Astron. Astrophys. **14,** 1221 (2014).

[102] L. Anderson *et al.* (BOSS), Mon. Not. R. Astron. Soc. **441,** 24 (2014).

[103] T. Delubac *et al.* (BOSS Collaboration), Astron. Astrophys. **574,** A59 (2015).

[104] M. Moresco, Mon. Not. R. Astron. Soc. **450,** L16 (2015).

[105] M. Moresco, L. Pozzetti, A. Cimatti, R. Jimenez, C. Maraston, L. Verde, D. Thomas, A. Citro, R. Tojeiro, and D. Wilkinson, J. Cosmol. Astropart. Phys. 05 (2016) 014.

[106] A. L. Ratsimbazafy, S. I. Loubser, S. M. Crawford, C. M. Cress, B. A. Bassett, R. C. Nichol, and P. Väisänen, Mon. Not. R. Astron. Soc. **467,** 3239 (2017).

[107] N. Borghi, M. Moresco, and A. Cimatti, Astrophys. J. Lett. **928,** L4 (2022).

[108] K. Jiao, N. Borghi, M. Moresco, and T.-J. Zhang, arXiv:2205.05701.

[109] L. Huang, Z. Huang, H. Zhou, and Z. Li, Sci. China Phys. Mech. Astron. **65,** 239512 (2022).

[110] Y.-S. Song and W. J. Percival, J. Cosmol. Astropart. Phys. 10 (2009) 004.

[111] M. Davis, A. Nusser, K. Masters, C. Springob, J. P. Huchra, and G. Lemson, Mon. Not. R. Astron. Soc. **413,** 2906 (2011).

[112] M. J. Hudson and S. J. Turnbull, Astrophys. J. Lett. **751,** L30 (2012).

[113] S. J. Turnbull, M. J. Hudson, H. A. Feldman, M. Hicken, R. P. Kirshner, and R. Watkins, Mon. Not. R. Astron. Soc. **420,** 447 (2012).

[114] C. Blake *et al.*, Mon. Not. R. Astron. Soc. **436,** 3089 (2013).

[115] T. Okumura *et al.*, Publ. Astron. Soc. Jpn. **68,** 38 (2016).

[116] D. Huterer, D. Shafer, D. Scolnic, and F. Schmidt, J. Cosmol. Astropart. Phys. 05 (2017) 015.

[117] A. Pezzotta *et al.*, Astron. Astrophys. **604,** A33 (2017).

[118] F. Qin, C. Howlett, and L. Staveley-Smith, Mon. Not. R. Astron. Soc. **487,** 5235 (2019).

[119] F. Avila, A. Bernui, E. de Carvalho, and C. P. Novaes, Mon. Not. R. Astron. Soc. **505,** 3404 (2021).

[120] E. Macaulay, I. K. Wehus, and H. K. Eriksen, Phys. Rev. Lett. **111,** 161301 (2013).

[121] C. Blake *et al.*, Astron. Astrophys. **642,** A158 (2020).

[122] S. de la Torre *et al.*, Astron. Astrophys. **608,** A44 (2017).

[123] A. Moradinezhad Dizgah and R. Durrer, J. Cosmol. Astropart. Phys. 09 (2016) 035.

[124] B. Ghosh and R. Durrer, J. Cosmol. Astropart. Phys. 06 (2019) 010.

[125] P. Astier *et al.* (SNLS Collaboration), Astron. Astrophys. **447,** 31 (2006).

[126] R. Laureijs *et al.* (EUCLID Collaboration), arXiv:1110.3193.

[127] D. Sobral-Blanco and C. Bonvin, Phys. Rev. D **104,** 063516 (2021).

[128] I. Tutusaus, D. Sobral-Blanco, and C. Bonvin, arXiv:2209.08987.